Untitled

LU MU

2020/12/5

Abstract

Introduction

Many factors can influence a person's choice of voting when it comes to the federal election and the information can be used in many ways. This poses an interesting question — what factors influence a person's choice to vote for the Liberal party or the Conservative Party? And what information or observation we can get by comparing the two results of people's choice of voting for these two parties.

In this report, we decided to further investigate this question by deeply analyzing the different factors that influence a person's choice to vote liberal or PC. Using the 2019 Canadian election results (phone survey) data set, we began by identifying various variables that we thought could influence our analysis. We decided to inspect the following variables:

- satisfaction of the federal government under Justin Trudeau
- feel about Justin Trudeau

- gender
- age
- province living in

To conduct the analysis, we decided to use logistic regression. In our study, variables such as satisfaction of the federal government under Justin Trudeau, feel about Justin Trudeau, and age are categorical variables; to accommodate these variables, we used dummy variables to incorporate them in the logistic regression.

After looking at the results from our logistic regression and using a significance level of 0.05, we inspected the p-values of different variables to determine if the variables significantly influence the probability a person votes for liberal or the PC. The rate of feel about Justin Trudeau was determined to be a significant factor; age except for those in their 25-44 was determined to be influential; Provinces and Female were also determined to be substantial.

The results from our analysis could be useful in many different ways. Information about factors that influence a person's choice to vote the Liberal Party or the Conservative Party can be used for information related to tell the party thier protenial voters and predict their election results.

In this report, we will begin by discussing the data set and then model our results.

Data

This analysis uses the data collected by the 2019 canadian federal election phone survey.

Table 1

```
survey_data <- read.csv("C:/Users/lumuc/Desktop/final 304/cleaned_data.csv") #read in the data
survey_data <- survey_data %>% #Select variables of interest
  select(q20,
         q11,
         q13,
         q3,
         q2,
         q4
         )

Table1 <- head(survey_data) %>%
  kable(align = "c",
        format = "simple",
        col.names = c("q20", "q11", "q13", "q3",
                      "q2", "q4"),
        ) %>%
  kable_styling(font_size = 5)
```

```
## Warning in kable_styling(., font_size = 5): Please specify format in
## kable. kableExtra can customize either HTML or LaTeX outputs. See https://
## haozhu233.github.io/kableExtra/ for details.
```

```
Table1
```

q20	q11	q13	q3	q2	q4
good	1	(2) Fairly satisfied	(1) Male	25-44	(5) Quebec
good	0	(2) Fairly satisfied	(1) Male	18-24	(5) Quebec
bad	0	(4) Not satisfied at all	(1) Male	25-44	(5) Quebec
bad	0	(4) Not satisfied at all	(1) Male	65+	(5) Quebec
bad	0	(4) Not satisfied at all	(1) Male	18-24	(5) Quebec
good	0	(2) Fairly satisfied	(1) Male	18-24	(5) Quebec

Model

The model that we are using is logistic regression. We are interested in how variables such as satisfaction of the federal government under Justin Tredeau, feel about Justin Tredeau, gender, age and province living in affect a person's decision to vote for liberal party. The response variable that we are interested in is whether a person has dicided to vote for liberal, which is a categorical variable with two levels: 0 represents not voting liberal, and 1 means voting liberal. The first explanatory variable is age. This is the variable that we want to be specific because the specific age instead of age group provides a lot more information than age groups...

Logistic regression allows us the simplicity of analyzing the categorical response variable while incorporating both numerical and categorical explanatory variables. Due to the large sample size and the nature of CES, the sample represents the population well for a rough estimate. If a more precise estimate is necessary, we can simply post-stratify on top of the logistic regression model results.

Logistic regression estimates $\beta_0 ... \beta_k$ in the following equation:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

where p is the probability of event A that we are interested in, $\beta_0$is the intercept, $x_1 ... x_K$ are our variables of interest and $\beta_1 ... \beta_k$ are parameters for each of these variables. Based on the result, we are able to estimate p for a particular case given all the variables.

In our case, we want to estimate $\beta_{satisfactionofgovernment}, \beta_{feelaboutJT}, \beta_{gender}, \beta_{age}, \beta_{province}$ in:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_{satisfactionofgovernment} x_{satisfactionofgovernment} + \beta_{feelaboutJT} x_{feelaboutJT} + \beta_{gender} x_{gender} + \beta_{age} x_{age} + \beta_{provi}$$

where p is the probability that a person votes liberal party, β_0 is the intercept, $\beta_{satisfactionofgovernment}$ is the parameter for the satisfaction of the federal government under Justin Tredeau, $\beta_{feelaboutJT}$ is the parameter for how people feel about Justin Tredeau, β_{gender} is the parameter for gender and β_{age} is the parameter for age and $\beta_{province}$ is the parameter for province people living in.

We use `glm()` from `stats` package in `R` to fir the model to our data. We use `as.factor()` to incorporate dummy variables for all the categorical variables: family income, education level and the type of population center. For each categorical variable with n levels, we need n-1 dummy variables to fully study its influence on our response variable.

The dummy variables setting ups are stated in table 3, 4 and 5 in the appendix.

Results

Table 2 summaries our model results:

```
#creating the summary table
my_log_res <- data.frame(
  variable = c("intercept","q13(2) Fairly satisfied", "q13(3) Not very satisfied ", "q13(4) Not satisfi
  estimate = c(-0.44676, -0.85429, -2.50036, -3.93136, 2.08188, 1.13344, 2.83283, -0.28391, 0.27460, 0.4
  pvalue = c("0.34184", "0.00398", "1.63e-14","< 2e-16","< 2e-16","2.01e-08","< 2e-16","0.01593","0.221
)
kable(my_log_res, caption = "Summary of Losgistic Estimates",
      label = "Data Source: GSS2017")
```

In an equation, this means

$$log(\frac{\hat{p}}{1-\hat{p}}) = \hat{\beta}_0 + \hat{\beta}_{fairlysatisfied} x_{fairlysatisfied} + \hat{\beta}_{Notverysatisfied} x_{Notverysatisfied} + \hat{\beta}_{Notsatisfiedatall} x_{Notsatisfiedatall} + \hat{\beta}_{good} x_{goo}$$

$$+ \hat{\beta}_{ok} x_{ok} + \hat{\beta}_{perfect} x_{perfect} + \hat{\beta}_{Female} x_{Female} + \hat{\beta}_{25-44} x_{25-44} + \hat{\beta}_{45-64} x_{45-64} + \hat{\beta}_{65+} x_{65+}$$

$$+ \hat{\beta}_{BC} x_{BC} + \hat{\beta}_{PEI} x_{PEI} + \hat{\beta}_{NS} x_{NS} + \hat{\beta}_{NB} x_{NB} + \hat{\beta}_{Quebec} x_{Quebec} + \hat{\beta}_{Ontario} x_{Ontario}$$
$$+ \hat{\beta}_{Manitoba} x_{Manitoba} + \hat{\beta}_{Saskatchewan} x_{Saskatchewan} + \hat{\beta}_{Alberta} x_{Alberta}$$

Notice that we have incorporated our categorical variables using dummy variables. "fairlysatisfied", "Notverysatisfied", "Notsatisfiedatall" represent the four different satisfaction to the federal government categories; "good", "ok" and "perfect" describe the four attitudes towards Justin Tredeau ; "25-44" and "45-64" and "65+" represent the four age groups. "BC","PEI","NS","NB","Quebec","Ontario","Manitoba","Saskatchewan", "Alberta" represent 10 provinces that people who are living in British Columbia, Prince Edward Island,

Table 2: Summary of Losgistic Estimates

variable	estimate	pvalue
intercept	-0.44676	0.34184
q13(2) Fairly satisfied	-0.85429	0.00398
q13(3) Not very satisfied	-2.50036	1.63e-14
q13(4) Not satisfied at all	-3.93136	$<$ 2e-16
q20good	2.08188	$<$ 2e-16
q20ok	1.13344	2.01e-08
q20perfect	2.83283	$<$ 2e-16
q3(2) Female	-0.28391	0.01593
q225-44	0.27460	0.22178
q245-64	0.46157	0.03850
q265+	0.59622	0.01105
q4(10) British Columbia	-0.51735	0.07172
q4(2) Prince Edward Island	-0.21058	0.55844
q4(3) Nova Scotia	-0.42614	0.22493
q4(4) New Brunswick	-0.42943	0.23468
q4(5) Quebec	-0.12663	0.66053
q4(6) Ontario	0.17115	0.54715
q4(7) Manitoba	-0.09502	0.78475
q4(8) Saskatchewan	-0.77834	0.04019
q4(9) Alberta	-0.52966	0.16239

Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan and Alberta respectively. "Female" represent the two categories for gender.

The model predicts the following result for the probability a person votes for the Liberal Party p, rounded to three decimal places. Since log is an one to one function with p, we say the change on $log(\frac{\hat{p}}{1-\hat{p}})$ is isomorphic to any change on p, the probalbility that a person in Canada votes for the Liberal Party.

$$log(\frac{\hat{p}}{1-\hat{p}}) = -0.45 - 0.85x_{fairlysatisfied} - 2.50x_{Notverysatisfied} - 3.93x_{Notsatisfiedatall} + 2.08x_{good} + 1.13x_{ok} + 2.83x_{perfect}$$

$$-0.28x_{Female} + 0.27x_{25-44} + 0.46x_{45-64} + 0.60x_{65+} - 0.52x_{BC} - 0.21x_{PEI} - 0.426x_{NS} - 0.429x_{NB} - 0.127x_{Quebec}$$

$$+0.171x_{Ontario} - 0.095x_{Manitoba} - 0.778x_{Saskatchewan} - 0.529x_{Alberta}$$

The interpretation of the dummy variables' prediction results is comparing it to a certain level that is not in the equation above. A $\hat{\beta}_{25-44} = 0.275$ does not mean the coefficient for having an age in the interval 25-44 is 0.275. It represents the difference of influence (on vote for liberal party decision) between having an age is 25-44 and having an age is 18-24. 0.275 indicates that having an age from 25 to 44 is more likely to vote for the Liberal than having an age is 18-24 if it is influential at all depending on the p-value.

Using $\alpha = 0.05, H_0 : \hat{\beta} = 0, \hat{\beta} \neq 0$, the p values indicate weak evidence that having a age in 25-44 age group affects a person's probability of voting for liberal party. At the same time, it is evident that other age groups do influence the probability. A p-value of $0.228 > 0.05$ indicates weak evidence to reject H_0; therefore, we cannot say staying in 25-44 age group influences a person's probability of voting for the liberal. However, staying in 45-64 age group and 65+ age group have influences on a person's probability of voting for the Liberal since p-value of 45-64 is $0.038 < 0.05$ and p-value of 65+ is $0.011 < 0.05$ in order to reject the null

hypothesis. We can also conclude that having older age is more likely to vote for the Liberal Party because the coefficient of 65+ which is 0.59 is larger than the coefficient of 45-64 which is 0.275.

There is no evidence that living in province like British Columbia, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba and Alberta have influences on p because their p values are bigger than 0.05. There is evidence that living in Saskatchewan variables influences p but its p-value is 0.040 which is very close to 0.05 but we don't consider it because p-value 0.04 means there is 50 percent probability of incorrectly rejecting a true null hypothesis and all other province shows no evidence of connection.

Having a higher satisfaction towards the federal government performance under Justin Trudeau is more likely to vote for the Liberal Party in 2019 federal election by comparing the coefficient of fairly satisfied which is -0.85, not very satisfied which is -2.50 and not satisfied at all which is -3.93. Same for the "how people feel about Justin Trudeau" variable. The higher rates that people give to Justin Trudeau, the more likely they will vote for the Liberal party 2019 whose leader is still him by comparing the coefficient of ok which is 1.13, good which is 2.08 and perfect which is 2.83. The lower the satisfaction towards the federal government or the lower rate for Justin Trudeau, the less chance of a person voting for the Liberal Party.

As we know, five parties had representatives elected to the federal parliament in the 2019 election: the Liberal Party who currently form the government, the Conservative Party who are the Official Opposition, the New Democratic Party, the Bloc Québécois, and the Green Party of Canada. The most two dominant party are the Liberal Party and the Conservative Party. It makes me think that will the result still the same if we keep the five variables unchanged, and investigate what factors influence a person's choice to vote for the Conservative Party?

Table 2

```
survey_data_2 <- read.csv("C:/Users/lumuc/Desktop/final 304/cleaned_data_2.csv") #read in the data
survey_data_2 <- survey_data_2 %>% #Select variables of interest
  select(q11,
         q20,
         q13,
         q3,
         q2,
         q4)

table2 <- head(survey_data_2) %>%
  kable(align = "c",
        format = "simple",
        col.names = c("q20", "q11", "q13", "q3",
                      "q2", "q4"),
        ) %>%
  kable_styling(font_size = 5)
```

```
## Warning in kable_styling(., font_size = 5): Please specify format in
## kable. kableExtra can customize either HTML or LaTeX outputs. See https://
## haozhu233.github.io/kableExtra/ for details.
```

```
table2
```

q20	q11	q13	q3	q2	q4
0	good	(2) Fairly satisfied	(1) Male	25-44	(5) Quebec
0	good	(2) Fairly satisfied	(1) Male	18-24	(5) Quebec

q20	q11	q13	q3	q2	q4
0	bad	(4) Not satisfied at all	(1) Male	25-44	(5) Quebec
0	bad	(4) Not satisfied at all	(1) Male	65+	(5) Quebec
0	bad	(4) Not satisfied at all	(1) Male	18-24	(5) Quebec
0	good	(2) Fairly satisfied	(1) Male	18-24	(5) Quebec

##Model 2 Logistic regression estimates $\beta_0...\beta_k$ in the following equation:

$$log(\frac{k}{1-k}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

where k is the probability of event A that we are interested in, $\beta_0$is the intercept, $x_1...x_K$ are our variables of interest and $\beta_1...\beta_k$ are parameters for each of these variables. Based on the result, we are able to estimate p for a particular case given all the variables.

In our case, we want to estimate $\beta_{satisfactionofgovernment}, \beta_{feelaboutJT}, \beta_{gender}, \beta_{age}, \beta_{province}$ in:

$$log(\frac{p}{1-p}) = \beta_0+\beta_{satisfactionofgovernment}x_{satisfactionofgovernment}+\beta_{feelaboutJT}x_{feelaboutJT}+\beta_{gender}x_{gender}+\beta_{age}x_{age}+\beta_{provi}$$

where k is the probability that a person votes the Conservative party, β_0 is the intercept, $\beta_{satisfactionofgovernment}$ is the parameter for the satisfaction of the federal government under Justin Tredeau, $\beta_{feelaboutJT}$ is the parameter for how people feel about Justin Tredeau, β_{gender} is the parameter for gender and β_{age} is the parameter for age and $\beta_{province}$ is the parameter for province people living in.

We use `glm()` from `stats` package in `R` to fir the model to our data. We use `as.factor()` to incorporate dummy variables for all the categorical variables: family income, education level and the type of population center. For each categorical variable with n levels, we need n-1 dummy variables to fully study its influence on our response variable.

The dummy variables setting ups are stated in table 3, 4 and 5 in the appendix.

Results 2

Table 2 summaries our model results:

```
#creating the summary table
my_log_res_2 <- data.frame(
  variable = c("intercept","q13(2) Fairly satisfied", "q13(3) Not very satisfied ", "q13(4) Not satisfi
  estimate = c(-16.0060, 14.3344, 15.5322, 16.3945, -2.0863, -1.0598, -2.6234, -0.4391, 0.2575, 0.6220,
  pvalue = c("0.958898", "0.963187", "0.960114","0.957901","< 2e-16 ***","1.18e-11 ***","4.91e-09 ***",
)
kable(my_log_res_2, caption = "Summary of Losgistic Estimates",
      label = "Data Source: GSS2017")
```

In this equation, this means

$$log(\frac{\hat{k}}{1-\hat{k}}) = \hat{\beta_0}+\hat{\beta}_{fairlysatisfied}x_{fairlysatisfied}+\hat{\beta}_{Notverysatisfied}x_{Notverysatisfied}+\hat{\beta}_{Notsatisfiedatall}x_{Notsatisfiedatall}+\hat{\beta}_{good}x_{go}$$

$$+\hat{\beta}_{ok}x_{ok} + \hat{\beta}_{perfect}x_{perfect} + \hat{\beta}_{Female}x_{Female} + \hat{\beta}_{25-44}x_{25-44} + \hat{\beta}_{45-64}x_{45-64} + \hat{\beta}_{65+}x_{65+}$$

$$+\hat{\beta}_{BC}x_{BC} + \hat{\beta}_{PEI}x_{PEI} + \hat{\beta}_{NS}x_{NS} + \hat{\beta}_{NB}x_{NB} + \hat{\beta}_{Quebec}x_{Quebec} + \hat{\beta}_{Ontario}x_{Ontario}$$

Table 4: Summary of Losgistic Estimates

variable	estimate	pvalue
intercept	-16.0060	0.958898
q13(2) Fairly satisfied	14.3344	0.963187
q13(3) Not very satisfied	15.5322	0.960114
q13(4) Not satisfied at all	16.3945	0.957901
q20good	-2.0863	< 2e-16 ***
q20ok	-1.0598	1.18e-11 ***
q20perfect	-2.6234	4.91e-09 ***
q3(2) Female	-0.4391	9.96e-05 ***
q225-44	0.2575	0.251542
q245-64	0.6220	0.004698 **
q265+	0.8330	0.000325 ***
q4(10) British Columbia	-0.0586	0.839638
q4(2) Prince Edward Island	0.5375	0.149051
q4(3) Nova Scotia	0.4372	0.234645
q4(4) New Brunswick	0.8404	0.022204 *
q4(5) Quebec	-0.5310	0.078600 .
q4(6) Ontario	0.3143	0.280948
q4(7) Manitoba	0.8074	0.014025 *
q4(8) Saskatchewan	1.1135	0.000763 ***
q4(9) Alberta	1.6632	8.16e-07 ***

$$+\hat{\beta}_{Manitoba}x_{Manitoba} + \hat{\beta}_{Saskatchewan}x_{Saskatchewan} + \hat{\beta}_{Alberta}x_{Alberta}$$

Notice that we have incorporated our categorical variables using dummy variables. "fairlysatisfied", "Notverysatisfied", "Notsatisfiedatall" represent the four different satisfaction to the federal government categories; "good", "ok" and "perfect" describe the four attitudes towards Justin Tredeau ; "25-44" and "45-64" and "65+" represent the four age groups. "BC","PEI","NS","NB","Quebec","Ontario","Manitoba","Saskatchewan", "Alberta" represent 10 provinces that people who are living in British Columbia, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan and Alberta respectively. "Female" represent the two categories for gender.

The model predicts the following result for the probability a person votes for the Conservative Party k, rounded to three decimal places. Since log is an one to one function with k, we say the change on $log(\frac{\hat{p}}{1-\hat{p}})$ is isomorphic to any change on k, the probalbility that a person in Canada votes for the Conservative Party.

$$log(\frac{\hat{k}}{1-\hat{k}}) = -16.006+14.334x_{fairlysatisfied}+15.532x_{Notverysatisfied}+16.395x_{Notsatisfiedatall}-2.086x_{good}-1.059x_{ok}-2.623x_p$$

$$-0.439x_{Female}+0.257x_{25-44}+0.622x_{45-64}+0.833x_{65+}-0.059x_{BC}+0.538x_{PEI}+0.437x_{NS}+0.840x_{NB}-0.531x_{Quebec}$$

$$+0.314x_{Ontario} + 0.314x_{Manitoba} + 1.114x_{Saskatchewan} + 1.663x_{Alberta}$$

There is strong evidence that living in province like New Brunswick, Manitoba, Saskatchewan and Alberta have influences on p because their p values much less than 0.05. By comparing the coefficient of these provinces, living in Alberta provinces has the highest chance of voting for the Conservative Party with the coefficient of 1.6632. Followed by Saskatchewan, New Brunswick and Manitoba with the coefficients of 1.114,

0.840 and 0.807. There is weak evidence that living in provinces like British Columbia, Prince Edward Island, Nova Scotia, Quebec and Ontario have influences p because their p-value is smaller than 0.05.

The four different attitudes which are very satisfied, fairly satisfied, not very satisfied and not satisfied at all towards the federal government do not show evidences to affect the probability of a person voting for the Conservative Party because their p-values are much larger than 0.05.

However, the variable how people feel about Justin Trudeau does show a significant influence on the probability of a person voting for the Conservative Party. Having a bad rate for Justin Trudeau has the highest chance of voting for the Conservative with the coefficient of o while as having a perfect rate for Justin Trudeau has the least chance of support the Conservative with a coefficient of -2.623.

Having a age in the 45-64 age group and 65+ age group have influence on a person of voting for the Conservative Party since their p-value is much less than 0.05. With the growth of ages, people tend to vote for the Conservative Party with the increase of coefficients.

Female are less likely to vote for the Conservative Party compared to male because the coefficient of female is -0.439.

Discussion

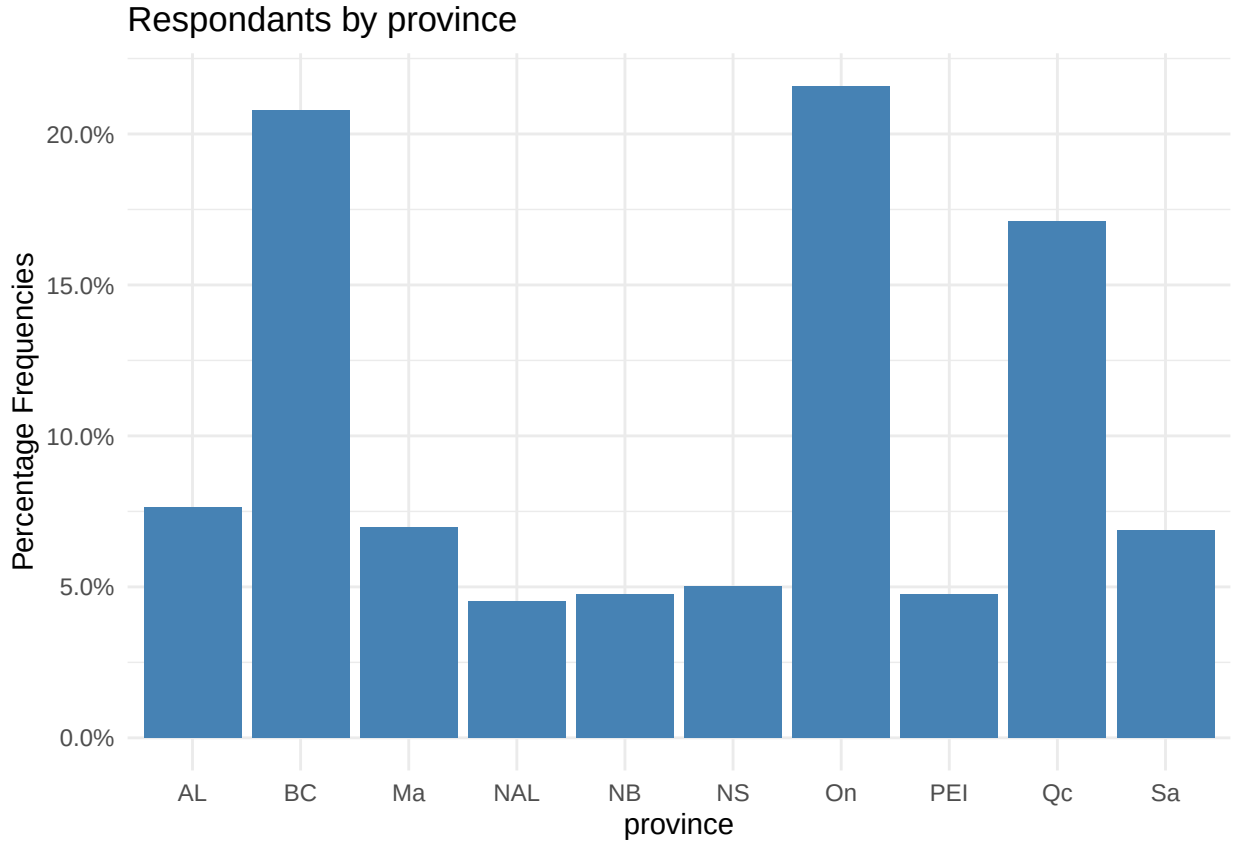
The 2019 federal election phone survey data has shown that there are significant effects on a person's choice to vote for the Liberal Party or the Conservative Party by satisfaction of the federal government under Justin Trudeau, feel about Justin Trudeau, gender, age, province living in. This data can be useful for two parties to identify their protential supporters and also urge the federal government do a better service for people.

Firstly, People who live in New Brunswick, Manitoba, Saskatchewan and Alberta province are more likely to vote for the Conservative party compared to other provinces. The reason that people live in these four provinces prefer to vote PC is the following: For Alberta, the Conservative party has taken Alberta almost entirely back to blue amid frustrations over pipeline construction and federal oil and gas policy, as the Liberal party lost all of the gains it made in Alberta in 2015. For New Brunswick, some districts are royal to the Conservative Party and their people are vote for PC by tradition. In its 102-year history of the southern New Brunswick district, it elected a Liberal only twice. For Manitoba, Liberals is not a presence in Manitoba. Here's why. Manitoba is a two-party system, with the NDP and the Conservatives regularly trading power in government. They tend to be "big tent" when it comes to policy ideas. Both parties at times straddle the centrist space left by having no competitive Liberal party. Because of the relative weakness of the Liberals, many liberal-minded individuals with political aspirations have either gone to the NDP or the Conservatives. There is a swing to the Conservatives began in the 2006 federal election, when the party captured 49 per cent of the vote in Saskatchewan, its best performance in more than 40 years since Diefenbaker had a hometown advantage. Ever since then, Andrew Scheer, another leader from Saskatchewan, makes the Conservatives great again by his leadership. These are the reason why the Conservative Party are dominant in these four provinces in the federal election.

However, it is not the case for liberal. Our result does not show that the province where people live in has a connection with the probability of voting for the Liberal. Even if living in Saskatchewan has a p-value of 0.04019, we still can not say that living in Saskatchewan has a significant influence on the probability of a person voting for the Liberal Party because its p-value is too close to 0.05.

Secondly, female are less likely to vote for the Conservative Party and the Liberal Party compared to male. It is a interesting observation because many people thought gender will not have influence on the probability of which party woman voting for when it comes to the federal election. But the fact is female are knowing their right and even running as candidates to pursue their political career in the Parliament. What party would woman vote for in the election if they do not support the Liberal and the Conservatives? The answer is NDP(New Democrats Party). NDP has the highest percentage of woman running as candidates among other four major parties. The 2011 election saw 452 women candidates out of a total 1587 people running (28.5%) and NDP has 39.9% of woman candidates and this percentage is the highest; This represented a record high proportion of candidates in Canadian federal elections. Why does woman like to vote for the NDP and

run as candidates in this party? Working life keeps getting harder for Canadians. People, especially young people and woman, are having a hard time finding secure, full-time jobs with benefits because they are more vulnerable compared to strong adult male. The growth of low-paid, insecure work is the result of decades of decisions by Liberal and Conservative governments – from scrapping the federal minimum wage in the 1990s to telling Canadians to "get used to" short-term, precarious work. New Democrats know that good jobs that treat people fairly make a real difference to Canadian families. We're committed to making sure that hardworking people get the fair wages and good working conditions they deserve. Therefore, many woman choose to vote for the NDP instead of the Liberal party or the Conservative Party. ### Figure 1

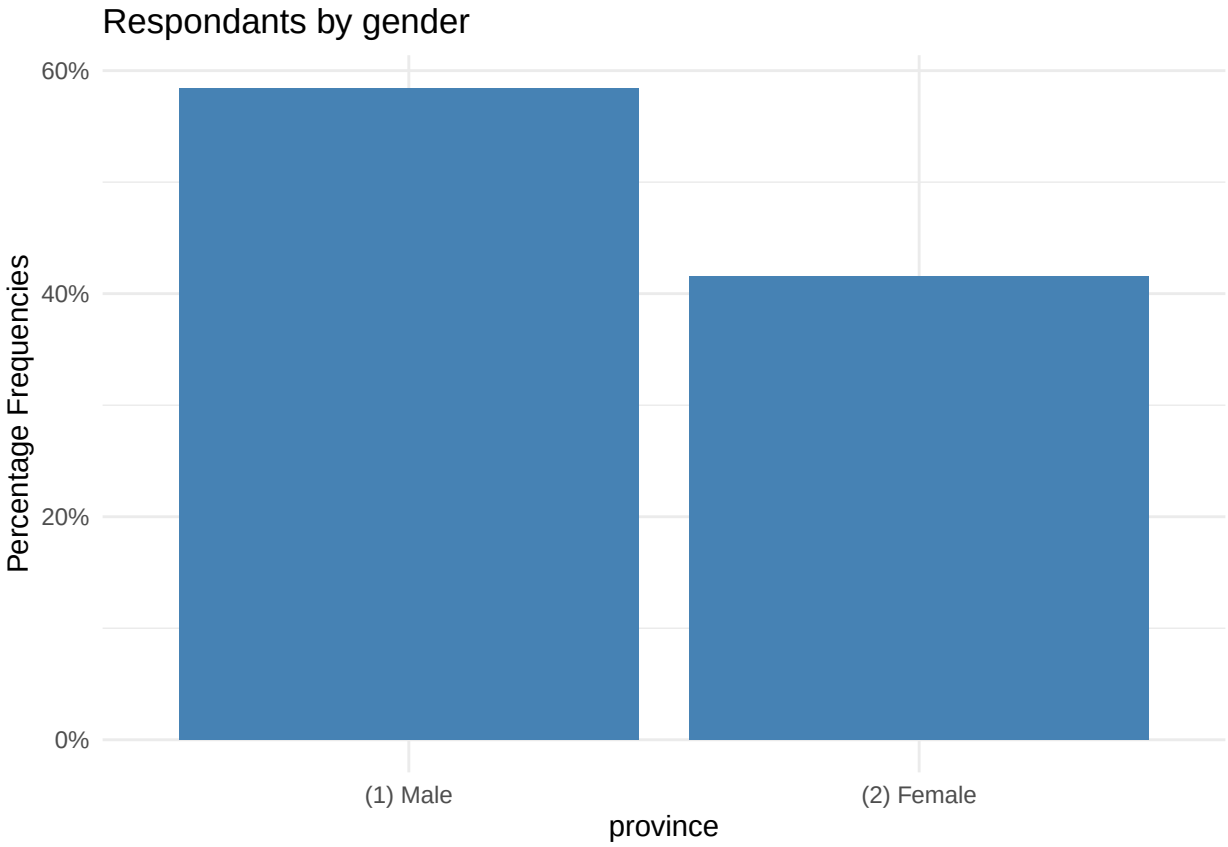


Province/ Territory	Percentage of combined provincial populations
British Columbia	13.31%
Alberta	11.00%
Saskatchewan	3.08%
Manitoba	3.64%
Ontario	38.91%
Quebec	23.22%
New Brunswick	2.20%
Nova Scotia	2.75%
Prince Edward Island	0.42%
Newfoundland and Labrador	1.49%
Provincial total	**100%**
Yukon	
Northwest Territories	
Nunavut	
Total	

Figure 1.B Canadian Female Population by Age 2014

Figure 1 displays the general breakdown of respondents province where they live. A chart which can show each province's respondent percentage account for the all respondents from any recent year federal election will be the best option to make comparsion. But we can not find it on the Internet. Instead, in Figure 1B, we took a table from election Canada website showing Percentage of combined provincial populations by assuming all the people can vote. As we can see, we can not match the percentage of each province in the first graph into the second's and can not find any similar pattern. The small sample size causes this error because our sample size is about 3000 and the actual number of respondents will be much larger than 3000.

Figure 2



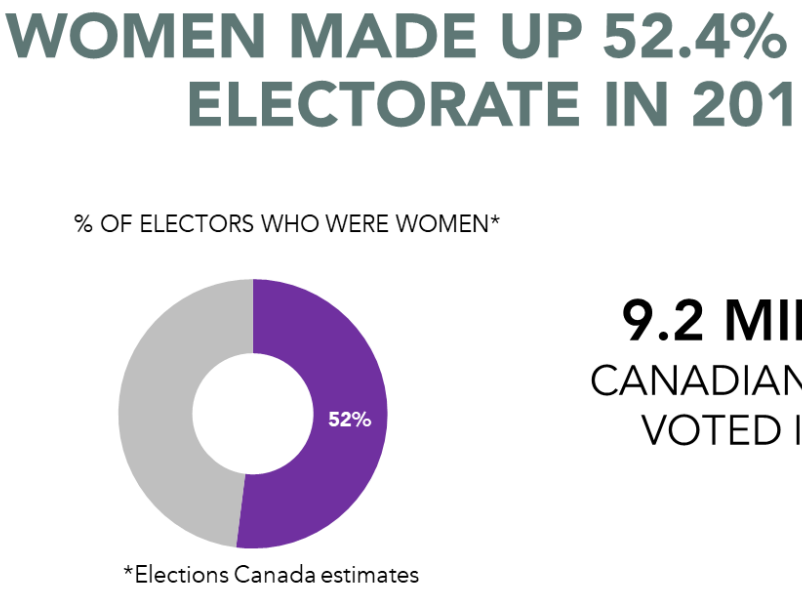


Figure 2.B Canadian Female Population

Figure 2 displays the general breakdown of respondents by gender. We can observe that male accounts for 60 percent among the respondents while as female accounts for about 40 percent of all the respondents.

Figure 2.B was taken from google image. We can see that women made up 52.4 percent of the federal election in 2015. Again, this difference of women voters in two elections could causes by the limited sample size in the phone survey. If we take a good amount of the sample size, the result of Figure 2 will be close to Figure 2.B.

##reference https://www.thestar.com/edmonton/2019/10/21/live-2019-federal-election-results-in-alberta.html https://www.cbc.ca/news/canada/new-brunswick/federal-election-new-brunswick-results-1.5329417 https://theconversation.com/manitobas-pragmatic-conservatism-may-contain-lessons-for-andrew-scheer-123671 https://www.cbc.ca/news/canada/saskatchewan/grenier-saskatchewan-political-dynaties-1.5747048 https://www.ndp.ca/economy https://www.google.com/search?q=how+many+male+and+female+vote+in+the+canadian+election&tbm=isch&ved=2ahUKEwj5xLWmvb_tAhWFBd8KHY4QB4wQ2-cCegQIABAA&oq=how+many+male+and+female+vote+in+the+canadian+election&gs_lcp=CgNpbWcQAzoECCMQJzo
wNg1_wDaBFwAHgAgAFZiAH_JJIBAjY5mAEAoAEBqgELZ3dzLXdpei1pbWfAAQE&sclient=img&ei=PwTQX7mpDIWL_AaOoZzgCA&bih=888&biw=1920&safe=strict#imgrc=40GQLb2cBEXLtM